Combining Polarizable Embedding with the Frenkel exciton model

Applications to absorption spectra with overlapping solute-solvent bands

Stendevad, Julie; Kongsted, Jacob; Svendsen, Casper Steinmann

Published in:
Theoretical Chemistry Accounts

DOI (link to publication from Publisher):
10.26434/chemrxiv.7504697
10.1007/s00214-019-2430-5

Publication date:
2019

Document Version
Early version, also known as pre-print

Combining Polarizable Embedding with the Frenkel exciton model: Applications to absorption spectra with overlapping solute-solvent bands

Julie Stendevad,[†] Jacob Kongsted,[†] and Casper Steinmann[*,‡]

[†]*Department of Physics, Chemistry and Pharmacy, University of Southern Denmark, Campusvej 55, DK–5230 Odense M, Denmark*

[‡]*Department of Chemistry and Bioscience, Aalborg University, Fredrik Bajers Vej 7H, DK–9220 Aalborg, Denmark*

E-mail: css@bio.aau.dk

Abstract

Modeling of spectral properties of extended chemical systems, such as the case of a solute in a solvent, is often performed based on so-called hybrid models in which only part of the complete system is given a quantum chemical description. The remaining part of the system is represented by an embedding potential treating the environment either by a discrete or continuum model. In order to successfully make use of minimally sized quantum chemical regions, the embedding potential should represent the environment as authentic as possible. Here, the importance of exactly such an accurate description of the embedding potential is investigated by comparing the performance of the Polarizable Embedding scheme against larger sized full quantum mechanical calculations. Our main conclusion is that as long as the solute and solvent do not overlap in

their absorption spectra, the Polarizable Embedding approach shows results consistent with full quantum chemical calculations. For partly overlapping absorption spectra the Polarizable Embedding approach can furthermore successfully be expanded within a Frenkel exciton approach based on only economical monomeric quantum chemical calculations. Thus, by extending the Polarizable Embedding scheme to the exciton picture it is possible to cover computations of the whole absorption spectrum and still reduce the computational cost compared to costly cluster calculations.

1 Introduction

In recent years, a lot of effort has been devoted to development and benchmarking of computational procedures aimed at calculating molecular response properties of chromophores in their natural environments such as in a solvent, a protein or another bio-structure. Because molecular response properties are coupled to the electronic structure of the molecule in question, such calculations require at least a partial quantum mechanical treatment, which in many cases is challenging due to the generally large size of the considered molecular systems and the associated computational cost of such calculations.

The continued concerted development of novel efficient algorithms and computer architectures has made it possible to perform quantum chemistry (QC) calculations on very big systems as highlighted recently in Ref.[1] Unfortunately, the procedure of simply increasing the size of the system treated with quantum mechanics (QM) still suffers from relative high computational cost especially when studying bio-molecular systems,[2] where effects of nuclear dynamics can play a crucial role. Usually such dynamics is handled by coupling the quantum mechanical method to e.g. molecular dynamics. In addition to this, problems with methods rooted in density functional theory (DFT) have been observed when increasing the size of the molecular system,[3] despite of DFT being the most used method for treating big molecular systems. Such problems usually show up in finite cluster representations of the system considered and leads, in the case of solute-solvent systems, to artificially edge effects.[4]

Hence, alternative methods to large cluster QM calculations seem to be well justified.

One typical approach to overcome the computational scaling when considering QM calculations on big systems is to apply hybrid quantum-classical methods, where the system of interest is divided into subsystems.[5–7] Here, the central part of the system (e.g. a chromophore) is treated with QM while the surroundings are treated as either a dielectric polarizable medium[8] or with a molecular mechanics (MM) based method, which allows much larger systems to be investigated. The clear advantage of the latter approach is that the structural details of the environment are preserved. The balance between computational cost versus adequate precision necessitates new models to be developed to imitate the surroundings in a fashion where accuracy and computational cost are balanced at the appropriate level depending on the questions to be answered.

The polarizable embedding (PE) model is a fragment based quantum-classical approach tuned to compute spectroscopic properties.[9] For a recent perspective on the background for the Polarizable Embedding method we refer to Ref.[10] In the PE method, atoms in the MM region are typically assigned distributed multipole moments up to quadrupoles to represent the static charge distribution of the environment. In addition polarizabilities are added to describe environmental polarization effects. These polarizabilities are usually approximated to be frequency independent, however, frequency dependent polarizabilities can also be used.[11] The polarizabilities give rise to an explicit description of polarization in the environment which is determined in a self-consistent manner, leading to a fully coupled description of polarization between the QM and the MM regions. The PE model is generally able to provide results that are consistent with full QM calculations, but benefits from significantly reduced computational cost as exemplified in a recent study of the predicted electronic absorption spectrum of the Green Fluorescent Protein.[12] It has been applied in the computation of optical properties over a wide range of target systems: from solute-solvent systems[9,13] to more complex setups with chromophores embedded in large biological systems as proteins[14,15] or nucleic acids.[16]

Recently, the PE model has also been formulated and applied in the context of the Frenkel exciton Hamiltonian[17] in order to compute excitation energies and electronic couplings between chromophores in multi-chromophoric proteins, where a given chromophore is coupled to other chromophores embedded in the protein through excitation energy transfer.[18] The use of the exciton model allows for a straightforward computation of coupled optical properties, and constructing the Frenkel Hamiltonian from properties derived form PE calculations represents an appealing procedure, especially for systems where the absorption spectrum of the chromophore is overlapping with the absorption spectrum of the surroundings.

In this work we compare different approaches to compute absorption properties of two model chromophores (two thioester analogs of the photo-active yellow protein (PYP) chromophore) in different solvent environments. For this we will rely on the standard PE approach and here investigate the requirements regarding the size of the QM system in order to reach results comparable to full QM calculations. Next, we will investigate the possibility of reducing the size of the QM region by coupling the PE approach to the exciton picture. For a coherent presentation we will begin the next section by shortly review the theoretical background of the PE approach as well as its coupling to the Frenkel picture.

2 Theoretical Background

Introducing the effect of the solvent environment through the polarizable embedding scheme allows for mutual polarization between the core QM region and the surrounding MM region. In the PE-DFT formalism this response is included through the effective Kohn-Sham operator given by:[9]

$$\hat{f}^{\text{eff}} = \hat{f}^{\text{KS}} + \hat{v}^{\text{PE}}. \tag{1}$$

The Kohn-Sham operator, $\hat{f}^{\text{KS}}$, includes all internal interactions in the isolated QM region, whereas the PE operator, $\hat{v}^{\text{PE}}$, describes the pertubation on the QM region exerted by the

environment, i.e. the embedding potential. The PE operator has two overall contributions[19]

$$\hat{v}^{\mathrm{PE}} = \hat{v}^{\mathrm{es}} + \hat{v}^{\mathrm{ind}}. \tag{2}$$

The electrostatic operator, $\hat{v}^{\mathrm{es}}$, represents the potential arising from the permanent multipoles in the surrounding region and is given by

$$\hat{v}^{\mathrm{es}} = \sum_{s=1}^{S} \sum_{|k|=0}^{K} \frac{(-1)^{|k|}}{k!} M_s^{(k)} V_{s,el}^{(k)} \tag{3}$$

Here, $M_s^{(k)}$ is the kth order multipole moment at site s, and the operator, $V_{s,el}^{(k)}$, provides the kth order derivative of the electronic electric potential at site s.[19] k is a three-dimensional multi-index notation of multipoles implying that $k = 0$ is the notation for a monopole (point charge), $k = 1$ is a dipole and so forth.

The second term in eq 2 is the induction operator, $\hat{v}^{\mathrm{ind}}$, which includes the environmental polarization here treated at the level of induced dipoles:[9]

$$\hat{v}^{\mathrm{ind}} = -\sum_{s=1}^{S} \boldsymbol{\mu}_s^{\mathrm{ind}} \left(\mathbf{F}_{tot}\right) \mathbf{F}_{s,el} \tag{4}$$

Here, $\mathbf{F}_{s,el}$, is the electronic electric field vector, whereas $\boldsymbol{\mu}_s^{ind}\left(\mathbf{F}_{tot}\right)$ represents the induced dipoles effected by the total electric field at the specific polarizable site. The induced dipoles are determined self-consistently leading to mutual polarization between the QM and MM region. To enable calculations of excited states, as well as general response properties, the PE model has been formulated within the concepts of quantum mechanical response theory.[9,20]Thus, this extension of PE provides direct access to spectroscopic properties in embedding calculations.

Recently, the PE model was further extended to the case of multi-chromophoric systems by formulating PE within the Frenkel exciton picture.[17,18] In this picture the excitation of M

molecules in a substance can be described trough the many-particle wave function defined as the Hartree product of the mth molecule in the pth excited state, ϕ_m^p, and the remaining molecules in their ground states

$$\Phi_m^p = \phi_m^p \prod_{\substack{n \neq m \\ m=1}}^{M} \phi_n^0 = \phi_1^0 \phi_2^0 \dots \phi_m^p \dots \phi_M^0. \tag{5}$$

The linear combination of all possible (single molecule) excited state Hartree products equals the kth exciton, Ψ^k,

$$\Psi^k = \sum_{m=1}^{M} \sum_{p} C_{m,p}^k \Phi_m^p. \tag{6}$$

The coefficients for the mth molecule in excited state p of the kth exciton, $C_{m,p}^k$, are found from diagonalization of the Hamiltonian matrix which is given by

$$\hat{H} = \sum_{m=1}^{M} \sum_{p} \epsilon_{m,p} \hat{a}_{m,p}^\dagger \hat{a}_{m,p} + \sum_{m}^{M} \sum_{n \neq m}^{M} \sum_{p,q} J_{mn}^{pq} (\hat{a}_{m,p}^\dagger \hat{a}_{n,q} + \hat{a}_{n,q}^\dagger \hat{a}_{m,p}). \tag{7}$$

Here, the first term includes the site energy of the mth molecule in the pth excited state. In the PE-DFT framework, the site energy is the excitation energy related to a single chromophore.[18] $\hat{a}_{m,p}^\dagger$ is a creation operator that creates the pth excited state on molecule m, and $\hat{a}_{m,p}$ is the corresponding annihilation operator. The second term in eq 7 describes the transfer of the excitation between molecule m and n. J_{mn}^{pq} is the exciton coupling between molecule m in the pth excited state and molecule n in the qth excited state, which, when calculated in presence of a polarizable environment, has two overall contributions[21]

$$J_{mn}^{pq} = J_{mn}^{pq\ (0)} + J_{mn}^{pq\ (1)}\ . \tag{8}$$

The direct interactions between molecule m and n are covered by the zeroth order term, $J_{mn}^{pq\ (0)}$, whereas the second term ($J_{mn}^{pq\ (1)}$) accounts for the response of the polarizable environment. Both the zeroth and the first order terms of J_{mn}^{pq} involve the transition densities

of molecule m and n. In this work we choose to evaluate J_{mn}^{pq} using a set of atomic partial charges fitted to reproduce the electrostatic potential of the transition density with a constraint to reproduce the transition dipole moment of the excitation. This atomic partial charge method is found to be highly efficient with respect to accuracy and computational costs.[18] Within this point charge representation of the transition density potential the first term of eq 8 becomes[22]

$$J_{mn}^{pq\ (0)} \approx \sum_{i\in m}\sum_{j\in n} \frac{q_i^T q_j^T}{|\mathbf{r}_i - \mathbf{r}_j|} \ . \tag{9}$$

The atomic partial charges, q_i^T and q_j^T located at points $\mathbf{r}_i$ and $\mathbf{r}_j$, here chosen to be atomic positions, are fitted to represent the potential of the transition density of molecule m and n, respectively. We assume that the distance between the molecules m and n is large enough to neglect electron exchange, and are therefore content with only including the Coulomb coupling presented in eq 9.

With an atomic partial charge model to represent the potential of the transition density the last term in eq 8 is written as[18]

$$J_{mn}^{(\mathbf{1})} \approx -\mathbf{F}\left(\left\{q_m^T\right\}\right)\mathbf{B}\mathbf{F}\left(\left\{q_n^T\right\}\right) \tag{10}$$

Here, $\mathbf{F}$ is transition electric field vectors at the polarizable sites calculated from the electronic transition density fitted charges located at the atoms in molecule m and n, respectively. The matrix $\mathbf{B}$ is the classical response matrix which is exactly the same matrix as used for determination of the electronic ground state induced dipole moments entering the embedding potential for determination of the electronic ground state of a given chromophore.

Having both the site energies and the exciton couplings available, the Frenkel Hamiltonian can be diagonalized, and the corresponding eigenvectors and eigenvalues (exciton transition energies) can be derived. From this, the wave function of the kth exciton (eq 6) and the exciton linear absorption stick spectrum is then obtained from the oscillator strength of the

kth excited state

$$f_k = \frac{2}{3}\Delta E^k_{m,p}\left|\boldsymbol{\mu}^T_k\right|^2, \tag{11}$$

with the exciton electric transition dipole moment, $\boldsymbol{\mu}^T_k$, for the kth exciton is given by

$$\boldsymbol{\mu}^T_k = \sum_{m=1}^{M}\sum_{p} C^k_{m,p}\boldsymbol{\mu}^T_{m,p}. \tag{12}$$

We note that the exciton coupling model is straightforward to apply to multi-chromophore systems since only one molecule is treated with QM at any one time thus keeping the computational costs down. As such, the exciton coupling model can be used to treat solute-solvent systems where the solvent molecules might become part of the exciton simply by treating some of the solvent molecules as chromophores and thus contributing to the explicit description of the delocalized excitation. This exciton procedure is compared to calculations where the same solvent molecules have been included directly in the QM part of the system. Thus, the question of whether or not the use of an exciton model allows for a (further) reduction of the size of the system treated using QM in cases where the solvent can be considered part of the excitation process, will be answered.

3 Methodology

Geometries of two thioester analogs of the PYP chromophore (shown in Figure 1) solvated in water were obtained from previous work by Milanese et al.[23] based on molecular dynamics (MD) simulations in TIP3P[24,25] water. The two analogs are, respectively, the neutral trans-thiophenyl-p-coumarate, (**1**), and the anionic form of it (**2**). 19 solvated structures of **1** and **2** in water were obtained from Ref. 23 and were used in this study without further modifications. Additionally, MD simulations of **1** and **2** in chloroform ($CHCl_3$) and in dimethyl sulfoxide (DMSO) were performed to test solvent effects on absorption spectra, as well as investigate the implication of applying the exciton model for systems where the

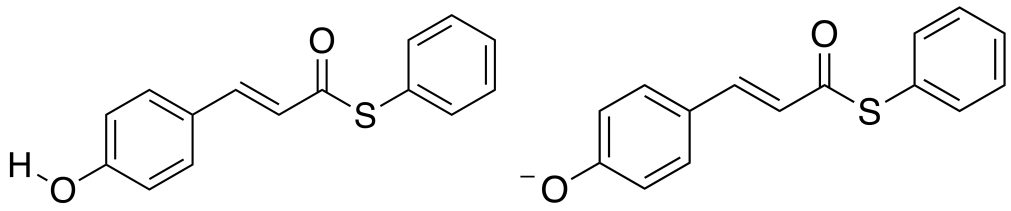

Figure 1: Illustration of solutes **1** (left) and **2** (right).

absorption spectra of the solute and solvent are partially overlapping. In the case of the solvent being $CHCl_3$ the MD simulations were performed using the Amber14 package.[26] The solutes were solvated in a cubic box with a radius from the solutes of 20 Å. The general Amber force field (GAFF)[27] was used for the chromophore analogs and $CHCl_3$ parameters from Cieplak et al.[28] were used to model the solvent. To equilibrate the systems, 1000 steps of initial minimization followed by 1000 steps of minimization of the solvent keeping the solute fixed with harmonic positional restraints were performed. After minimization, three MD equilibration steps were performed using a stepwise procedure starting by heating the system to 300 K over 20 ps with constant volume, after which the systems were allowed to equilibrate at 300 K for 10 ps. The last equilibration step consisted of a 200 ps MD at 300 K using periodic boundary conditions to assure that the system was completely equilibrated. After equilibration 10 ns of Langevin dynamics were performed at a temperature of 300 K and 1 atm pressure using a time step of 2 fs. Bonds involving hydrogens were constrained with SHAKE[29] in all MD runs. From the production run, snapshots were extracted from 1 ns every 0.5 ns for the full simulation yielding a total of 19 snapshots. For the one-photon absorption spectra, the solutes and the nearest 400 solvents were included.

For the simulations of the solutes in DMSO, the MD simulations were performed with Desmond.[30,31] Both **1** and **2** were solvated in a cubic box of DMSO with a radius of 10 Å. The OPLS3e[32] force field was applied for all molecules. The systems were relaxed using Desmonds standard relaxation protocol before production runs within the constant-pressure, constant-temperature ensemble using a pressure of 1 atm and a temperature of 300 K and using SHAKE. Like the simulations in $CHCl_3$, the production simulations were run for 10

ns during which snapshots were extracted every 0.5 ns to be used in the simulations of the spectra and only the solutes and the nearest 400 solvents were included in later simulations.

3.1 One-Photon Absorption Spectra

The one-photon UV/Vis absorption spectra presented below were obtained from PE-TDDFT calculation using the fully long-range corrected CAM-B3LYP[33] functional and the 6-31+G*[34,35] basis set. All PE-TDDFT calculations were performed in the Dalton program.[36,37] In all the presented spectra, the lowest lying 12 transition energies and oscillator strengths for all systems were used to compute absorption spectra broadened with Gaussian line shapes with a broadening factor of 0.2 eV.

The solvent molecules were described using different embedding potentials when computing the one-photon absorption spectra. In PE-type calculations, a hierarchy of models are usually introduced when describing the parameters entering Eqs. 3 and 4. The hierarchy is based on the order of the static electric moments (0 for charge, 1 for dipole and so on) or the polarizability tensors (0 for non-polarizable, 1 for isotropic and 2 for fully anisotropic dipole-dipole polarizabilities). With such a definition, water can be described by either TIP3P[38,39] (M0P0, i.e. a charge based model without polarizabilities), the WAT model[40] (M0P1, i.e. a charge based model with isotropic polarizabilities) or using a distributed multipole moment expansion to second order (quadrupoles) on atoms and anisotropic electric dipole-dipole polarizabilities also placed on atoms (M2P2). In this study, the M2P2 parameters were derived by utilizing the LoProp approach[41] employing the LoProp for Dalton script[42] based on integrals and response functions obtained also from the Dalton program package. M2P2 parameters are computed for each water molecule without reuse of any parameters. In the calculation of the one-photon absorption spectra of the solutes solvated in $CHCl_3$, the solvent parameters were the CL3[40] model (M0P1). Similarly, the embedding parameters for DMSO were the DMS[40] model (M0P1).

To quantify how a spectrum S compares with a reference spectrum S^{ref} the integral[43–45]

$$\delta_{\text{dev}}(S, S^{\text{ref}}) = \frac{\int_{\omega_0}^{\omega_1} |S(\omega) - S^{\text{ref}}(\omega)| \, \mathrm{d}\omega}{\int_{\omega_{\text{min}}}^{\omega_{\text{max}}} |S^{\text{ref}}(\omega)| \, \mathrm{d}\omega} \tag{13}$$

is evaluated using Simpon's rule in SciPy.[46] When two spectra are perfectly overlapping the deviation, $\delta_{\text{dev}}(S, S^{\text{ref}})$, evaluates to zero. The integral limits ω_0 and ω_1 are defined in this work by the frequency range of the presented spectra.

3.2 Systems

Because of the time-consuming nature of performing TD-DFT calculations, especially when also taking averaging over multiple snapshots into account, this work is focused on the effect of reducing the size of the QM region to a minimum whilst maintaining an accuracy comparable to the use of larger QM regions.

3.2.1 Water

When using water as a solvent, a reference system (**W**) of either solute **1** or **2** is constructed by including the 50 closest water molecules to either of the chromophores in the quantum region with the remaining water molecules treated by the TIP3P water model. To study the effect of demoting solvent molecules from the QM region to either one of the classical potentials described above, several smaller systems based on **W** are constructed. Three models, **W1**, **W2** and **W3** are constructed by replacing the 50 water molecules with TIP3P, WAT and M2P2, respectively. Additional models based on **W1**, **W2** and **W3** were also constructed where the water molecules thath are hydrogen-bonding to the solutes are included in the QM region. For all **W1**, **W2** and **W3** the remaining solvents surrounding the 50 specified solvents are still treated as TIP3P.

3.2.2 Chloroform

For systems with $CHCl_3$ as solvent a reference system (**C**) of either solute **1** or **2** was constructed from the first snapshot extracted from the MD simulation with the chromophore and the 12 closest $CHCl_3$ molecules included in the QM region. The remaining $CHCl_3$ molecules were treated with the CHL model. A single model was constructed where $CHCl_3$ in the QM region was replaced with the CHL model (**C1**). Also, a model where only $CHCl_3$ within 3.0 Å were included in the QM region was constructed (**C2**). For **C1** and **C2** the remaining $CHCl_3$ were still treated with the CHL model.

3.2.3 DMSO

For systems with DMSO as solvent, the reference system (**D**) was constructed from the first snapshot extracted from the MD simulation, again with only the 16 closest DMSO molecules included in the QM region. The remaining DMSO molecules were treated with the DMS model. Similarly to the $CHCl_3$ systems two models were subsequently generated (**D1** and **D2**) where, respectively, all DMSO solvents in the QM region are replaced with the DMS model or only DMSO within 3.0 Å were included in the QM region and the remaining solvent molecules were treated with the DMS model.

3.3 Exciton Absorption Spectra

One-photon absorption spectra were also computed using the exciton model presented in Section 2 at the level of theory presented in Section 3.1. All calculations were performed only on the **W3**, **C2** and **D2** systems, i.e. structures with only the solvent molecules closest to the chromophore included in the QM region. The exciton approach is tested for accuracy compared to the traditional cluster approach.

4 Results and Discussion

It has previously been shown that using an appropriate embedding scheme to describe the MM region can considerably reduce the size of the QM region and thus the computational time without compromising accuracy.[12,47] Any deficiencies in reproducing the spectra found in this first part of the study will be attempted to be corrected for using a bottom-up approach by including more solvent molecules in the QM region, or as shown in the last part of the section, though the exciton approch .

4.1 Small QM Region Compared with Reference

4.1.1 Water

The reference one-photon absorption spectrum computed for solutes **1** and **2** in water (Figure 2, black lines) is compared with the three approximate models using only the first snapshot from the MD simulations. This is done to compare the effect of replacing water molecules described by QM and demoting them to classical models as explained in the computational details section above. Focusing initially on solute **1** (Figure 2A), the effect of

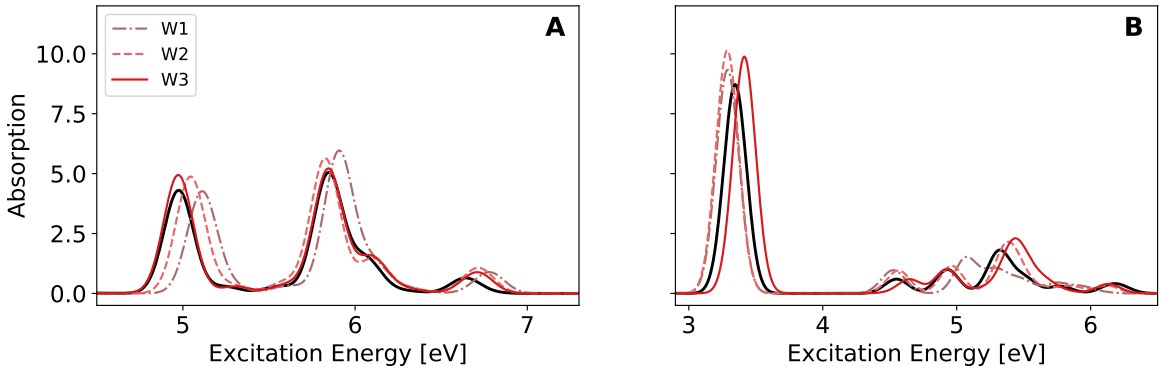

Figure 2: Absorption spectra computed with a minimal QM region for solutes **1** (left) and **2** (right) in water. The reference spectrum is shown in black.

replacing all 50 water molecules in the QM region with TIP3P water (**W1**) is observed to reproduce the overall features of the absorption spectrum with three peaks at 5.2, 5.9 and

6.8 eV, respectively. It is here noted that the spectrum is shifted 0.1–0.2 eV to the right for all peaks and that using TIP3P water fails to reproduce the shoulder observed on the central peak at around 6.2 eV. On the other hand, using either the polarizable WAT water model (**W2**) or the polarizable M2P2 model (**W3**), this shoulder is reproduced. Additionally, it is also observed that the peak positions are much better reproduced (0.0–0.1 eV), especially for the first two peaks. This indicates that mutual polarization effects between solute and solvents are the major contributor to reproduce these low-lying excitations. The improvement of using WAT or M2P2 over TIP3P for solute **1** is also evident when evaluating eq 13 where the computed deviations to the reference are 0.72, 0.45 and 0.16 for **W1**, **W2** and **W3**, respectively. The absorption spectrum for solute **2** (Figure 2B) shows only one major peak in the low energy area (3.3–3.4 eV) while several peaks of lower intensity are observed at higher energies (4.3–6.2 eV). Here, it is observed that all models reproduce the placement and height of the first peak. TIP3P (**W1**) shows several deficiencies with the peaks in the rest of the spectrum, whereas the two other models (**W2** and **W3**) reproduce the spectrum to a better extent. There is, however, an indication that for M2P2 there might be over-polarization effects because the major peak position is shifted 0.1 eV to a peak position of 3.5 eV. The latter is also evident when evaluating eq 13 where the observed deviations are 0.59, 0.52 and 0.63 for **W1**, **W2** and **W3**, respectively, i.e., overall slightly worse than for the neutral solute.

As expected the overall polarization effects are observed to be important when replacing QM waters with MM waters. However, the difference observed between the WAT and M2P2 water models is minor and for one-photon absorption properties not observed to be significant. The M2P2 model is computed for each water molecule individually and this adds both complexity and simulation time when performing simulations. As observed in earlier work,[40] hybrid models combing WAT with M2P2 have been observed to provide an excellent compromise keeping M2P2 waters near the QM region with the rest either being treated with WAT or TIP3P. Nonetheless, such hybrid models are not investigated in this

work.

4.1.2 Chloroform

Solvating solutes **1** and **2** in $CHCl_3$ gives the one-photon absorption spectra presented in Figure 3 in black. Here, comparison with only the polarizable CHL model is done for the first snapshot. As was observed for water in the previous section, the prediction of the first

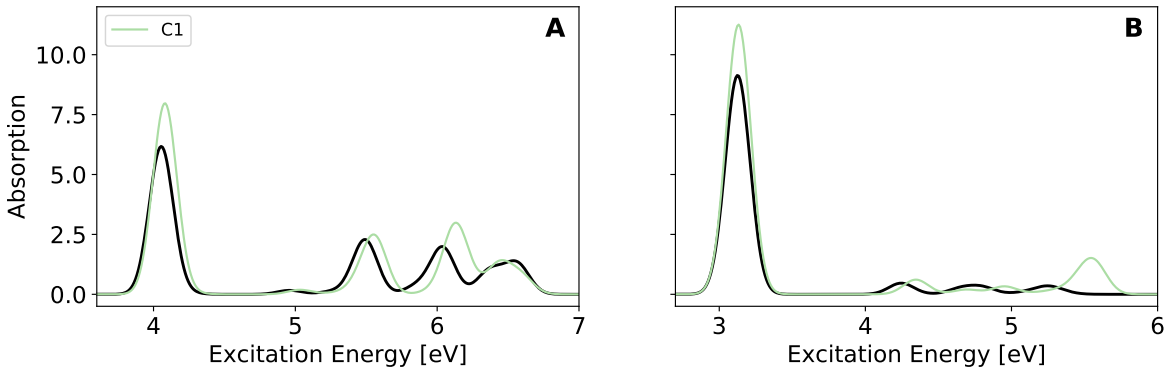

Figure 3: Absorption spectra using minimal QM region for solutes **1** (left) and **2** (right) in $CHCl_3$.

major peak, here observed at around 4.1 eV for solute **1** (Figure 3A) and 3.2 eV for solute **2** (Figure 3B), is reproduced with good accuracy using the CHL model with only the peak-height being overestimated somewhat. The high-energy region of the spectra predictions of peaks 2 (5.5 eV) and 4 (6.6 eV) are somewhat reproduced for solute **1** with the CHL model, whereas the third peak (6.0 eV) deviates by 0.2 eV and is predicted to be more intense than the reference calculation. A similar observation is made for solute **2** where all minor peaks are shifted 0.2–0.4 eV and the last predicted peak (5.5 eV) is also predicted to be 5 times more intense active than what the reference spectrum suggests. The deviations from the reference spectrum computed from eq 13 is 0.45 for both solutes.

4.1.3 Dimethyl Sulfoxide

In DMSO, the reference one-photon absorption spectra are presented in Figure 4 (black) for solutes **1** (Figure 4A) and **2** (Figure 4B). Replacing the quantum mechanically treated DMSO molecules with the polarizable DMS solvent model yields the blue spectra in Figure 4 For the neutral solute **1**, the obtained results are very much like the results obtained for both

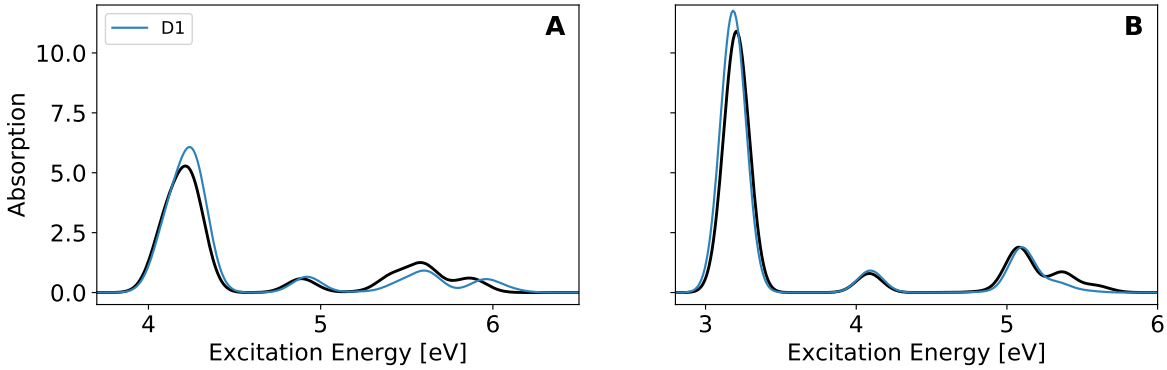

Figure 4: Absorption spectra using minimal QM region for solutes **1** (left) and **2** (right) in DMSO.

water and $CHCl_3$. The major peak at around 4.2 eV is reproduced with good accuracy ($<$ 0.1 eV for the position) and in the high-energy range, the peak positions are shifted around 0.1 eV to the right. The peak heights are, however, reproduced with good accuracy. For the anionic solute **2**, the major peak at 3.3 eV is reproduced with very good accuracy also in terms of peak height. Here, it is seen that only the 5+ eV range fails to reproduce the reference spectrum. The deviations from the reference spectrum computed from eq 13 is 0.27 and 0.24 for solutes **1** and **2**.

In summary, polarization effects are observed to be important for water where the non-polarizable TIP3P embedding potential is shown to have shortcomings in reproducing large QM/MM calculations. However, when addressing the same question looking at all three different solvents, using a polarizable model yields results which are more comparable to the reference calculation, especially in the low-energy region. For higher lying states there appears, however, still to be discrepancies which are most likely due to coupling between

the electronic states of the solute and the solvent and will be investigated in the following sections.

4.2 Large QM Region Compared with Reference

Based on the comparisons made above, the next step is to re-introduce solvent molecules into the QM region in order to remedy the discrepancies, particularly in the high-energy region of the computed spectra.

4.2.1 Water

For solutes **1** (Figure 5A) and **2** (Figure 5B), hydrogen bound water molecules were re-introduced into the QM region. For the first snapshot for each solute, this means that there are, respectively, 4 and 8 water molecules included in the QM region. For the negatively charged solute **2** the increased number of water molecules is due to more favorable interactions with the surrounding water molecules. Again, the remaining water molecules are

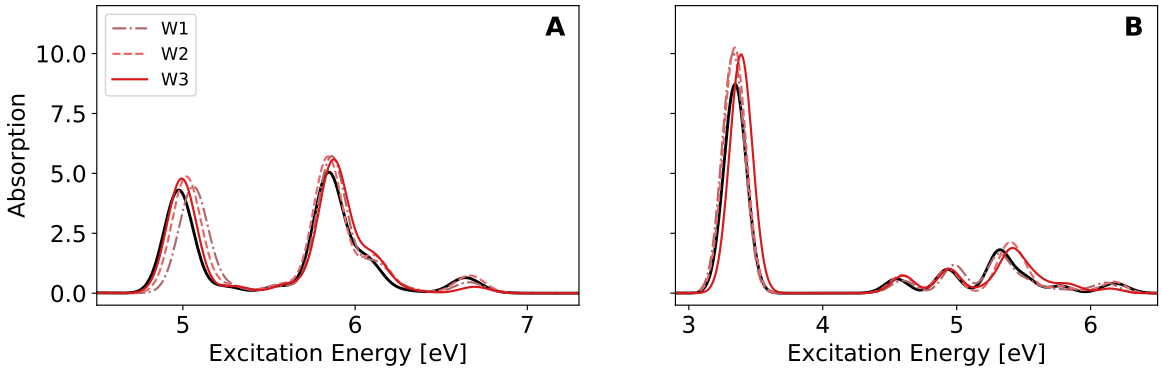

Figure 5: Absorption spectra computed with a large QM region for solutes **1** (left) and **2** (right) in water. The reference spectrum is shown in black.

treated with either TIP3P, WAT or M2P2 embedding potentials. Compared to the results obtained for the small QM region (Figure 2) there are definite improvements for solute **1** also when using TIP3P as an embedding potential. Specifically, progress is observed around the central peak at 5.9 eV where all three embedding potentials are now reproducing this

part of the spectrum as opposed to only the polarizable models in the previous section. This again confirms that the peak position and intensity are dominated by mutual polarization of the solutes and the solvents. Considering that even the peak around 6.7 eV is observed to improve for all three embedding potentials, it is suggested that the effect has character of energy transfer instead of being influenced by polarization effects. It is noted that the position of the peak around 5.0 eV is still poorly reproduced by the TIP3P potential. For solute **2**, similar conclusions can be made where an overall improvement across the entire spectrum is observed. This is valid for both the peak position (3.3 eV) in the low-energy regime and also for the higher energy excitations. However, at the peak around 5.3 eV a shift of 0.1 eV is observed for the polarizable models. This discrepancy is most likely due to over-polarization effects which could probably be ameliorated by including some form of damping when solving for the induced moments.[19]

When computing the deviations of the spectra to the reference spectrum for solute **1** one obtains 0.38, 0.27 and 0.24 for **W1**, **W2** and **W3**, respectively. This is, for **W1** and **W2** almost an improvement by a factor of two. However, because the spectrum for model **W3** with the small QM region was very good ($\delta_{\text{dev}} = 0.16$), an increase in deviation is observed for the large QM region to 0.24. Similar results are obtained for solute **2** where the improvements for all three water models are almost a factor of three for TIP3P (**W1**) down to a more modest improvement for the M2P2 potential (**W3**) of a factor 1.4.

4.2.2 Chloroform

Spectra computed for solutes **1** and **2** with all $CHCl_3$ molecules within 3.0 Å promoted to the QM region are presented in Figure 6A and Figure 6B, respectively. The major peak at 4.1 eV for solute **1** is reproduced with better accuracy than for the minimal QM region (Figure 3) with especially the peak height showing an improvement. However, the peaks at 6.1 eV and 6.4 eV show little to no improvement. For the anionic solute **2**, however, the high-energy peak that was observed at 5.5 eV in Figure 3 is absent and the spectrum is all

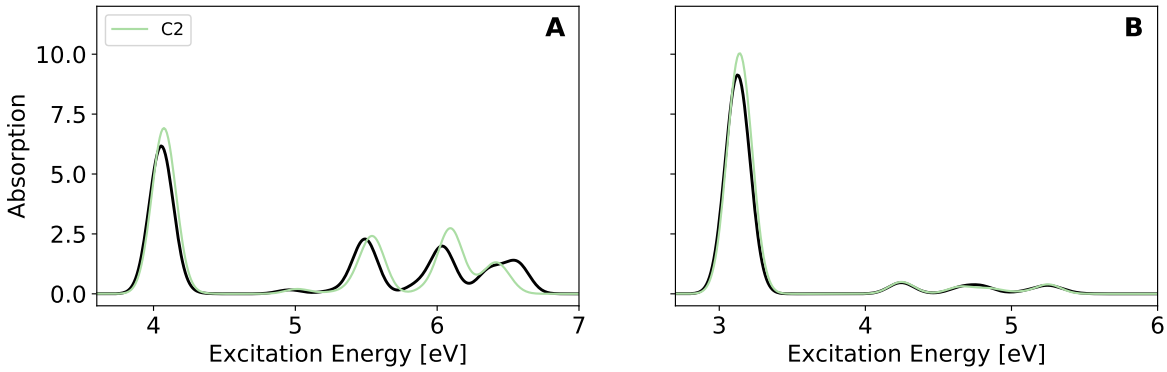

Figure 6: Absorption spectra computed with a large QM region for solutes **1** (left) and **2** (right) in $CHCl_3$. Reference spectrum shown in black.

over improved also at the major peak around 3.1 eV. When increasing the size of the QM region the computed spectra reproduce the reference spectra more accurately than for the small QM region. For solute **1** a modest improvement from 0.45 to 0.35 is observed, but solute **2** improves significantly from 0.45 to 0.15.

4.2.3 Dimethyl Sulfoxide

Spectra computed for solutes **1** and **2** with all DMSO molecules within 3.0 Å promoted to the QM region and the remaining DMSO molecules emloying the DMS solvent model are presented in Figure 7A and Figure 7B, respectively. For both solute **1** and **2** clear

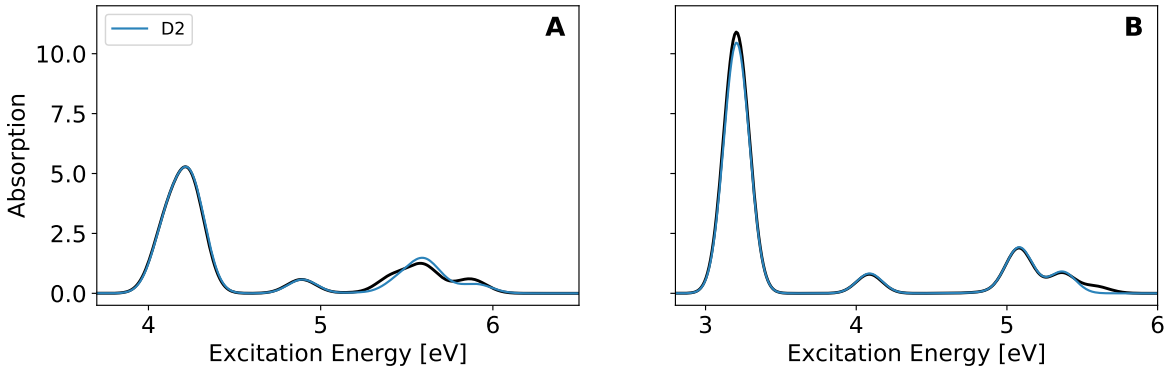

Figure 7: Absorption spectra computed with a large QM region for solutes **1** (left) and **2** (right) in DMSO. Reference spectrum shown in black.

improvements are observed when compared to the spectrum obtained for a minimal QM region (Figure 4). All major peaks are reproduced both in terms of position and peak height with the only minor disagreement observed around the peak at 5.6 eV for the neutral solute **1**. Compared to the reference spectrum this is considered to be negligible which is confirmed with deviations computed using eq 13 that are 0.07 and 0.05 for solutes **1** and **2**, respectively which are clear improvements from the small QM regions which were 0.27 and 0.24, respectively.

4.3 Exciton Based Large QM Region Compared with Reference

Based on the exciton model presented in Section 2, absorption spectra were computed on the same structures as was done in the previous section, but from the exciton transition energies and the exciton oscillator strength (eq 11). The resulting spectra using the exciton model for solutes **1** and **2** in water are presented in Figure 8. As was observed for the systems

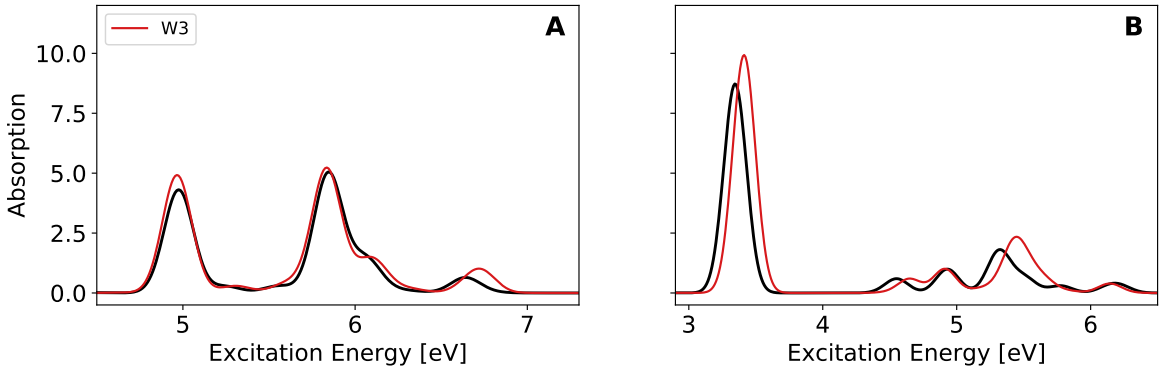

Figure 8: Absorption spectra of solutes **1** (left) and **2** (right) computed with the exciton model in water.

with the large QM region (Figure 5 for water), the use of an exciton model also improves the resulting spectra compared to only including the solute in the QM region (Figure 2). Because the mid- to high-energy regions are close to the excitation energies of water this is expected. The peak positions at 5.0 eV and 5.9 eV are well-reproduced for the neutral solute (Figure 8A) with a 0.1 eV shift observed for the high-energy peak at 6.7 eV. For

the anionic solute (Figure 8B) the low-lying peak at 3.3 eV is observed be to shifted by an equal amount (+0.1 eV) as observed in Figure 2B when using the M2P2 potential. This

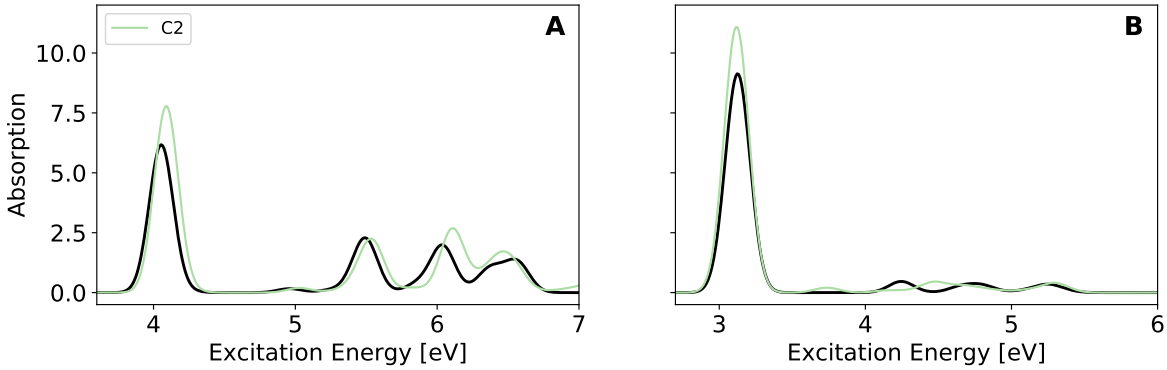

Figure 9: Absorption spectra of solutes **1** (left) and **2** (right) computed with the exciton model in $CHCl_3$.

could indicate, as stated before, an over-polarization that cannot be ameliorated by using an exciton model for this particular system. Computing the deviation from the reference spectrum (see Table 1) also confirms that the behavior is more like the small QM region with a deviation of 0.64. Either damping[19] of the induced moments or the WAT potential (as was observed in Figure 2B) could be employed for a better result. Similarly, both $CHCl_3$ (Figure 9) and DMSO (Figure 10) systems are observed to display an accuracy that lies between small and large QM regions presented above. Absorption spectra for solutes **1** and **2** in $CHCl_3$ are presented in Figure 10. Here, it is observed that for solute **1** the accuracy is more similar to the results obtained for the small QM region, emphasizing that other inter-molecular interactions than the solute-solvent Coulomb interaction are responsible for the considerable improvement observed for the large QM region (Figure 6). This is also confirmed when computing the deviations in spectra from the reference where 0.44 and 0.29 are obtained using eq 13. A similar conclusion is obtained for solute **2** in $CHCl_3$. For solute **1** in DMSO, there is an overall good agreement with the reference spectrum in both the low- and high- region of the spectrum. As was observed for water, however, solute **2** shows a minor shift (<0.1 eV) of the major peak at 3.2 eV which again must be contributed to

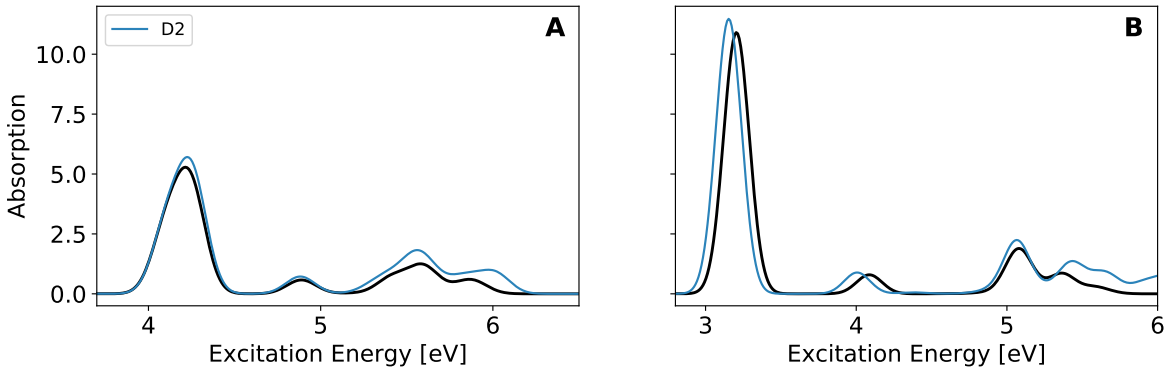

Figure 10: Absorption spectra of solutes **1** (left) and **2** (right) computed with the exciton model in DMSO.

a polarization effect. The remaining spectrum is represented with good accuracy both the small peak around 4.0 eV and the slightly larger peak at 5.1 eV. The deviations for solute **1** and **2** are, respectively, 0.26 and 0.56. The latter result is about a factor of two worse than purely using the small QM region (0.24) but the deviation stems mostly from the slight shift in the major absorption peak and the increase in absorption from 5.3 eV to 6.0 eV. In

Table 1: Computed deviations between a reference absorption spectrum, S^{ref}, of water and spectra for small, S^{small}, or large, S^{large}, QM regions or exciton based S^{exciton}.

		W1	**W2**	**W3**	**C2**	**D2**
1	$\delta_{\text{dev}}(S^{\text{small}}, S^{\text{ref}})$	0.72	0.45	0.16	0.45	0.27
	$\delta_{\text{dev}}(S^{\text{large}}, S^{\text{ref}})$	0.38	0.27	0.24	0.35	0.07
	$\delta_{\text{dev}}(S^{\text{exciton}}, S^{\text{ref}})$			0.21	0.44	0.26
2	$\delta_{\text{dev}}(S^{\text{small}}, S^{\text{ref}})$	0.59	0.52	0.63	0.45	0.24
	$\delta_{\text{dev}}(S^{\text{large}}, S^{\text{ref}})$	0.22	0.22	0.45	0.15	0.05
	$\delta_{\text{dev}}(S^{\text{exciton}}, S^{\text{ref}})$			0.64	0.28	0.56

conclusion we find that the exciton based model is able to reproduce a reference spectrum for all tested solvents with good accuracy by combining multiple chromophore calculations. The accuracy of the exciton model is, however, not exceeding the corresponding results based on the use of the large QM region. All computed deviations for the water spectra are summarized in Table 1.

4.4 Averaged Spectra

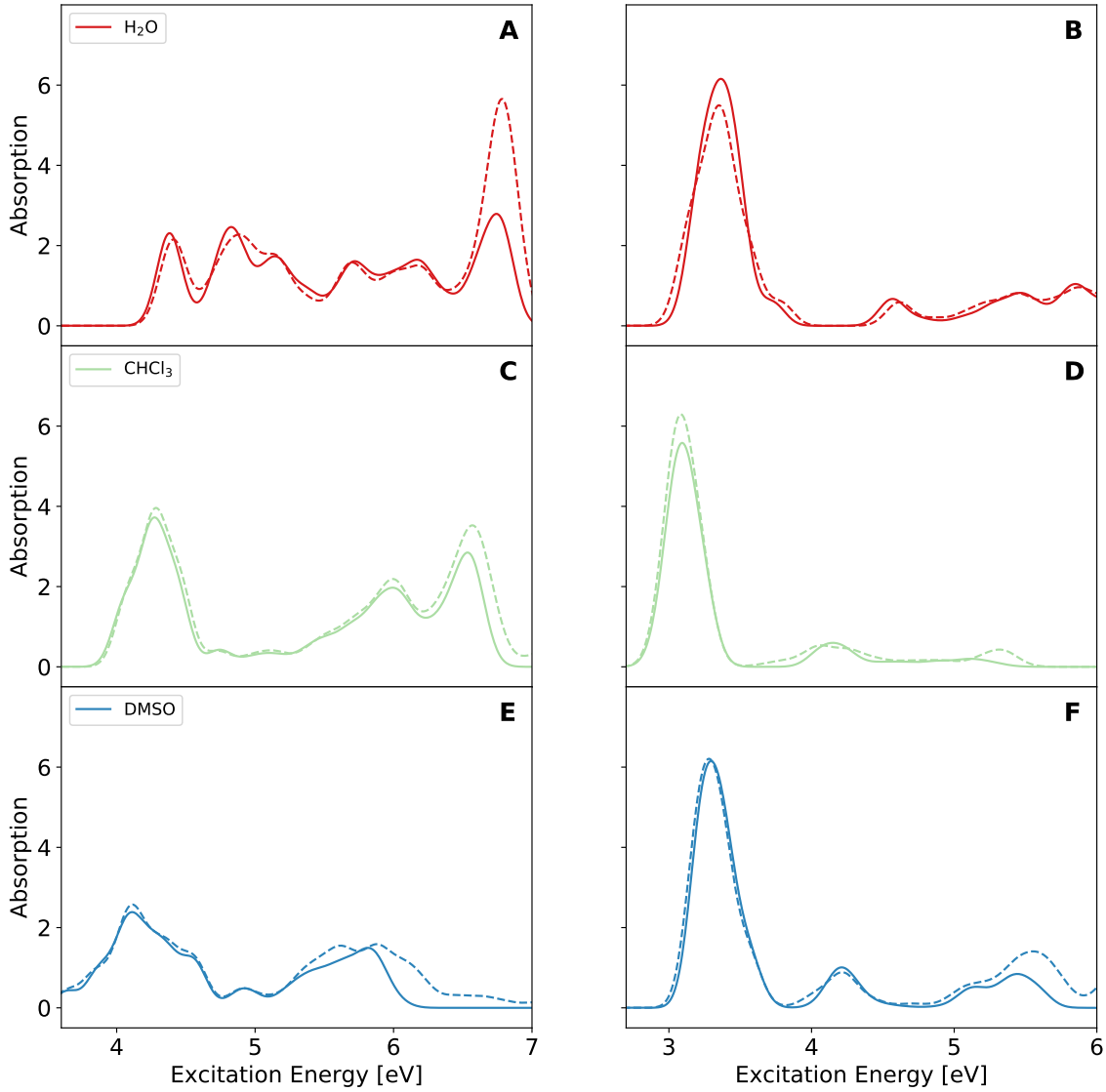

Figure 11: Averaged absorption spectra for solutes **1** (left) and **2** (right) in all tested solvents computed using PE-TDDFT cluster calculations (solid lines) and the exciton (dashed lines) model.

Based on the analysis above it is clear that adequate results, when compared to much larger reference calculations, can be obtained both with the traditional PE-TDDFT approach using large QM regions (models **W3**, **C2** and **D2**) or with the exciton approach.

Traditionally, averages over multiple snapshots are carried to potentially make comparisons with experimental results. In this section, averaged spectra computed with the PE-TDDFT approach are compared directly with averaged spectra from the exciton model. These results are presented in Figure 11 with solid lines (PE-TDDFT) or with dashed lines (exciton model) and it is noted that only polarizable embedding potentials are used: WAT for water, CHL for $CHCl_3$ and DMS for DMSO.

The major difference between the two solutes (**1** and **2**) in all three solvents is the very definitive major peak at 3.5 eV for water (Figure 11B), 3.1 eV for $CHCl_3$ (Figures 11D) and 3.5 eV for DMSO (Figure 11F) which is absent for the neutral solute shown in Figures 11A, 11C and 11E, respectively. For PE-TDDFT, the neutral solute shows no clear absorption peaks for water and only minor definitive peaks in the 4.0–4.4 eV section of the spectrum for DMSO and $CHCl_3$. The averaged spectra computed using the exciton-based model (dashed lines) are quite similar to the PE-TDDFT spectra where it is observed that the overall line shapes of the exciton spectra reproduce the spectra obtained with PE-TDDFT. For water, a high-energy peak at 6.8 eV is observed to be more pronounced with the exciton model in Figure 11A, and this is likely due to an increased interaction with the surrounding water molecules through the interaction presented in eq 8.

5 Conclusions

In this paper several approaches to compute the one-photon absorption spectrum of the neutral and anionic thioester analogs of the PYP chromophore solvated in either water, chloroform or dimethyl sulfoxide were investigated. All approaches made use of the Polarizable Embedding model potentially combined with the Frenkel exciton model to obtain absorption spectra. First, it was investigated how the different embedding schemes performed when only the chromophore was included in the QM region, i.e. a minimal QM region, and the remaining system was treated with classical models. Here, it was found that

representing the solvent using polarizable models in general resulted in spectra comparable to the reference spectrum. It was clear that the low-energy region of the spectrum was more accurately reproduced compared to the high-energy region. The non-polarizable TIP3P water model was on the contrary found have some problems reproducing the reference spectrum both in the low and high energy regions, which was clear when compared to the polarizable water models. Thus, use of minimal quantum regions should in such cases as considered here be performed in combination with the use of a polarizable model for description of the solvent.

Next, the QM region was expanded to include solvent molecules closest to the chromophore. For water, this meant that all hydrogen bonded water molecules were included, whereas for chloroform and dimethyl sulfoxide it was all solvent molecules within a radius of 3 Å from the chromophore. Overall, this expansion of the QM region corrected a considerable amount of the discrepancies observed for the minimal-sized QM regions, and particularly improvements in the high-energy region of the computed spectra were observed. However, some dissonance between the use of polarizable water models and the reference spectrum were observed for the anionic solute. The concern was especially one peak in the high energy region of the spectrum, which was ascribed to most likely be due to over-polarization effects. Lastly, the Frenkel exciton Hamiltonian was constructed by considering the solvent molecules from the large QM region. These results were compared to the result of the full QM cluster calculations. In general the exciton model reproduced the reference spectra for all three solvents with an accuracy comparable to be between the small and large QM calculations. This was observed as minor peak shifts in both the chloroform and the dimethyl sulfoxide (for the anionic chromophore) spectra, which most likely is due to over polarization effects. Thus, this study confirms that the expensive QM cluster calculations can be replaced with lower cost Polarizable Embedding exciton model calculations to a good approximation.

In the future, studies with full spectral overlap such as pure liquids will be investigated where the exciton model surely is applicable.

Acknowledgement

The authors would like to thank Joel Milanese for providing the structures of the water systems used in this work.

References

(1) Bowler, D.; Miyazaki, T. *Rep. Prog. Phys.* **2012**, *75*, 036503.

(2) Isborn, C. M.; Luehr, N.; Ufimtsev, I. S.; Martínez, T. J. *J. Chem. Theory Comput.* **2011**, *7*, 1814–1823.

(3) Jakobsen, S.; Kristensen, K.; Jensen, F. *J. Chem. Theory Comput.* **2013**, *9*, 3978–3985.

(4) Isborn, C. M.; Mar, B. D.; Curchod, B. F. E.; Tavernelli, I.; Martínez, T. J. *J. Phys. Chem. B* **2013**, *117*, 12189–12201.

(5) Warshel, A.; Levitt, M. *Journal of Molecular Biology* **1976**, *103*, 227 – 249.

(6) Field, M. J.; Bash, P. A.; Karplus, M. *Journal of Computational Chemistry* **1990**, *11*, 700–733.

(7) Lin, H.; Truhlar, D. G. *Theoretical Chemistry Accounts* **2007**, *117*, 185.

(8) Tomasi, J.; Mennucci, B.; Cammi, R. *Chemical Reviews* **2005**, *105*, 2999–3094.

(9) Olsen, J. M.; Aidas, K.; Kongsted, J. *J. Chem. Theory Comput.* **2010**, *6*, 3721–3734.

(10) List, N. H.; Olsen, J. M. H.; Kongsted, J. *Physical Chemistry Chemical Physics* **2016**, *18*, 20234–20250.

(11) Nørby, M. S.; Vahtras, O.; Norman, P.; Kongsted, J. *Mol. Phys.* **2016**, *115*, 39–47.

(12) Nåbo, L. J.; Olsen, J. M. H.; Martínez, T. J.; Kongsted, J. *J. Chem. Theory Comput.* **2017**, *13*, 6230–6236.

(13) Schwabe, T.; Olsen, J. M. H.; Sneskov, K.; Kongsted, J.; Christiansen, O. *J. Chem. Theory Comput.* **2011**, *7*, 2209–2217.

(14) List, N. H.; Olsen, J. M. H.; Kongsted, J. *Physical Chemistry Chemical Physics* **2016**, *18*, 20234–20250.

(15) Beerepoot, M. T.; Steindal, A. H.; Kongsted, J.; Brandsdal, B. O.; Frediani, L.; Ruud, K.; Olsen, J. M. H. *Physical Chemistry Chemical Physics* **2013**, *15*, 4735–4743.

(16) Nørby, M. S.; Steinmann, C.; Olsen, J. M. H.; Li, H.; Kongsted, J. *J. Chem. Theory Comput.* **2016**, *12*, 5050–5057.

(17) Frenkel, J. *Phys. Rev.* **1931**, *37*, 17–44.

(18) Steinmann, C.; Kongsted, J. *J. Chem. Theory Comput.* **2015**, *11*, 4283–4293.

(19) Steinmann, C.; Reinholdt, P.; Nørby, M. S.; Kongsted, J.; Olsen, J. M. H. *Int. J. Quantum Chem.* **2018**, *119*, e25717.

(20) Olsen, J. M. H.; Kongsted, J. *Advances in Quantum Chemistry*; Elsevier, 2011; Vol. 61; pp 107–143.

(21) Curutchet, C.; noz Losa, A. M.; Monti, S.; Kongsted, J.; Scholes, G. D.; Mennucci, B. *J. Chem. Theory Comput.* **2009**, *5*, 1838–1848.

(22) Madjet, M. E.; Abdurahman, A.; Renger, T. *J. Phys. Chem. B* **2006**, *110*, 17268–17281.

(23) Milanese, J. M.; Provorse, M. R.; Alameda, E.; Isborn, C. M. *J. Chem. Theory Comput.* **2017**, *13*, 2159–2171.

(24) Jorgensen, W. L. *J. Am. Chem. Soc.* **1981**, *103*, 335–340.

(25) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.

(26) Case, D. A.; Babin, V.; Berryman, J. T.; Betz, R. M.; Cai, Q.; Cerutti, D. S.; Cheatham, T. E.; Darden, T. A.; Duke, R. E.; Gohlke, H.; Goetz, A. W.; Gusarov, S.; Homeyer, N.; Janowski, P.; Kaus, J.; Kolossváry, I.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Paesani, F.; Roe, D. R.; Roitberg, A.; Sagui, C.; Salomon-Ferrer, R.; Seabra, G.; Simmerling, C. L.; Smith, W.; Swails, J.; Walker,; Wang, J.; Wolf, R. M.; Wu, X.; Kollman, P. A. Amber 14. University of California, San Francisco, 2014.

(27) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

(28) Cieplak, P.; Caldwell, J.; Kollman, P. *J. Comp. Chem.* **2001**, *22*, 1048–1057.

(29) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. *Journal of Computational Physics* **1977**, *23*, 327–341.

(30) Bowers, K. J.; Chow, D. E.; Xu, H.; Dror, R. O.; Eastwood, M. P.; Gregersen, B. A.; Klepeis, J. L.; Kolossvary, I.; Moraes, M. A.; Sacerdoti, F. D.; Salmon, J. K.; Shan, Y.; Shaw, D. E. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. ACM/IEEE SC 2006 Conference (SC06). 2006.

(31) Schrödinger Release 2018-4, Desmond Molecular Dynamics System, D. E. Shaw Research, New York, NY, 2018. Maestro-Desmond Interoperability Tools, Schrödinger, New York, NY. 2018.

(32) Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A. *J. Chem. Theory Comput.* **2015**, *12*, 281–296.

(33) Yanai, T.; Tew, D. P.; Handy, N. C. *Chem. Phys. Lett.* **2004**, *393*, 51–57.

(34) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257–2261.

(35) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213–222.

(36) Aidas, K.; Angeli, C.; Bak, K. L.; Bakken, V.; Bast, R.; Boman, L.; Christiansen, O.; Cimiraglia, R.; Coriani, S.; Dahle, P.; Dalskov, E. K.; Ekström, U.; Enevoldsen, T.; Eriksen, J. J.; Ettenhuber, P.; Fernández, B.; Ferrighi, L.; Fliegl, H.; Frediani, L.; Hald, K.; Halkier, A.; Hättig, C.; Heiberg, H.; Helgaker, T.; Hennum, A. C.; Hettema, H.; Hjertenæs, E.; Høst, S.; Høyvik, I.-M.; Iozzi, M. F.; Jansík, B.; Jensen, H. J. Aa.; Jonsson, D.; Jørgensen, P.; Kauczor, J.; Kirpekar, S.; Kjærgaard, T.; Klopper, W.; Knecht, S.; Kobayashi, R.; Koch, H.; Kongsted, J.; Krapp, A.; Kristensen, K.; Ligabue, A.; Lutnæs, O. B.; Melo, J. I.; Mikkelsen, K. V.; Myhre, R. H.; Neiss, C.; Nielsen, C. B.; Norman, P.; Olsen, J.; Olsen, J. M. H.; Osted, A.; Packer, M. J.; Pawlowski, F.; Pedersen, T. B.; Provasi, P. F.; Reine, S.; Rinkevicius, Z.; Ruden, T. A.; Ruud, K.; Rybkin, V. V.; Sałek, P.; Samson, C. C. M.; de Merás, A. S.; Saue, T.; Sauer, S. P. A.; Schimmelpfennig, B.; Sneskov, K.; Steindal, A. H.; Sylvester-Hvid, K. O.; Taylor, P. R.; Teale, A. M.; Tellgren, E. I.; Tew, D. P.; Thorvaldsen, A. J.; Thøgersen, L.; Vahtras, O.; Watson, M. A.; Wilson, D. J. D.; Ziolkowski, M.; Ågren, H. *WIREs Comput. Mol. Sci.* **2014**, *4*, 269–284.

(37) Dalton, a molecular electronic structure program, Release Dalton2016.3 (2016), see `http://www.daltonprogram.org`.

(38) Jorgensen, W. L. *J. Am. Chem. Soc.* **1981**, *103*, 335–340.

(39) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.

(40) Beerepoot, M. T. P.; Steindal, A. H.; List, N. H.; Kongsted, J.; Olsen, J. M. H. *J. Chem. Theory Comput.* **2016**, *12*, 1684–1695.

(41) Gagliardi, L.; Lindh, R.; Karlström, G. *J. Chem. Phys.* **2004**, *121*, 4494.

(42) Vahtras, O. LoProp for Dalton. 2014.

(43) Hopmann, K. H.; Ruud, K.; Pecul, M.; Kudelski, A.; Dračínský, M.; Bouř, P. *J. Phys. Chem. B* **2011**, *115*, 4128–4137.

(44) Štěpánek, P.; Bouř, P. *Phys. Chem. Chem. Phys.* **2014**, *16*, 20639–20649.

(45) Nørby, M. S.; Olsen, J. M. H.; Steinmann, C.; Kongsted, J. *Journal of Chemical Theory and Computation* **2017**, *13*, 4442–4451.

(46) Jones, E.; Oliphant, T.; Peterson, P.; Others, SciPy: Open source scientific tools for Python. 2001–2018; `http://www.scipy.org/`, [Online; accessed December 21st.].

(47) Steinmann, C.; Bratholm, L. A.; Olsen, J. M. H.; Kongsted, J. *J. Chem. Theory Comput.* **2017**, *13*, 525–536.